

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number  
**WO 02/08890 A2**

(51) International Patent Classification<sup>7</sup>: **G06F 9/00**

(21) International Application Number: **PCT/GB01/03282**

(22) International Filing Date: **20 July 2001 (20.07.2001)**

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:  
**0018042.2** **21 July 2000 (21.07.2000)** **GB**

(71) Applicant (for all designated States except US): **MONSELL EDM LTD.** [GB/GB]; Monsell House, Monsell Lane, Upton-on-Severn, Worcestershire WR8 0QN (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **LAFONTAINE, Seiner, Geoffrey, Thomas** [GB/GB]; Monsell House, Monsell Lane, Upton-on-Severn, Worcestershire WR8 0QN (GB).

(74) Agents: **PERKINS, Sarah** et al.; Stevens Hewlett & Perkins, Halton House, 20/23 Holborn, London, Greater London EC1N 2JD (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **METHOD OF AND SOFTWARE FOR RECORDAL AND VALIDATION OF CHANGES TO MARKUP LANGUAGE FILES**

(57) Abstract: A method of recording changes to a markup language file which employs application-defined tags. The changes are recorded in a delta file which is also a markup language file providing validation of the recorded changes against substantially the same markup language structure as that of the markup language file being changed. Where the original markup language file is an XML file with a DTD, a DTD can be created for the delta file which substantially follows the DTD of the original markup language file. Strict compliance of the data recorded in the delta file with the delta DTD provides validation of the changes with respect to the original XML file.

**BEST AVAILABLE COPY**

**WO 02/08890 A2**

## METHOD OF AND SOFTWARE FOR RECORDAL AND VALIDATION OF CHANGES TO MARKUP LANGUAGE FILES

5

The present invention relates to a method of managing changes to or between markup language files having application-defined markup tags and to software for performing the method such that the changes may be validated, tracked and recorded. More particularly the present invention is suited for use, but not exclusively, with XML documents/files.

XML (Extensible Markup Language) has been established as a standard for encoding information in electronic files for transfer between systems and across the world-wide-web (www). In particular, XML is used for the transfer of structured documents and data. Thus, XML is a set of rules, guidelines and conventions for designing text formats for data. Like HTML, XML is a markup language and so makes use of tags and attributes but while HTML specifies what each tag and attribute means XML uses the tags only to delimit pieces of data, and leaves the interpretation of the data completely to the application that uses it, hence the term application-defined tags. XML encodes, using these markup tags, a description of a document's storage layout and logical structure. Thus markup takes the form of start-tags, end-tags, empty-element tags, entity references, character references, comments, CDATA section delimiters, document type declarations and processing instructions. The markup tags are nested in a tree structure with the allowed structure for each tag, for example which tags may appear within a tag, formally defined in an element type declaration. Thus the structure of an XML file or document is specified by a set of element type declarations which is known as a Document Type Definition (DTD). The DTD is thus a grammar for the XML document. XML files may be large (tens of megabytes) and complex (over a thousand element type declarations within a DTD) especially where the XML file

contains CAD data. Hence, a file recording changes to an XML file can also tend to be large and complex. "Extensible Markup Language (XML) I-O W3C Recommendation 10 February 1998" (<http://www.w3.org/TR/REC-xml>) provides a much more detailed explanation of XML and its contents is  
5 incorporated herein by reference.

It is often necessary, for example, in software testing and change control support, to determine the exact differences, if any, between two markup language files and to encode these in some way preferably in a separate document that will be referred to herein as a delta file. Delta files  
10 are known in other areas of software development, however, traditionally files have been treated as a sequence of data records, typically lines of text. Therefore, differences of order or even 'white space' for example spaces, new lines, carriage return/line feed are considered significant. This can be particularly problematic where more complex data structures such  
15 as CAD systems are involved. Ideally, though, a method that identifies and records these differences should ignore apparent differences, for example the ordering of XML element attributes, which do not constitute semantic differences (see for example Canonical XML Version 1.0: <http://www.w3.org/TR/xml-c14n>, and the XML Information Set  
20 <http://www.w3.org/TR/xml-infoset>).

One approach to the analysis and recordal of differences between XML documents is the XML TreeDiff Update Language (<http://www.alphaworks.ibm.com/>). XML TreeDiff computes the differences between two XML documents by analysing the documents as DOM trees  
25 and identifying the differences as a sequence of tree editing operations (DOM is the acronym for Document Object Model). XML TreeDiff represents the differences between two XML files in a structure quite separate from that of the original XML files.

The present invention seeks to provide an improved method  
30 of and software for recording changes to markup language files that employ application-defined tags and that additionally enables such changes to be

validated. In the context of this document reference to a markup file being valid is reference to the contents of a markup file being well-formed and ideally complying with the constraints expressed in an associated set of element type declarations. Where a markup file has no associated set of  
5 element type declarations or where strict compliance with the document type definition of the markup file is not required, the markup file is deemed to have a document type definition where each element has a content model of ANY and the content of the markup file is validated by ensuring the content is well-formed.

10       Accordingly, the present invention provides a method of recordal of changes to a markup language file having application-defined tags wherein a delta file is created and changes to the contents of the markup language file are recorded in the delta file where the changes are validated against substantially the same markup language structure as the  
15 markup language file to be changed.

          Ideally, the delta file is created using the same markup language as the markup language file to be changed. Also, preferably, the delta file is assigned a plurality of element type declarations determined from the element type declarations of the markup language file, the  
20 differences between one or more of the element type declarations of the delta file and the element type declarations of the markup language file being restricted to permissible differences with respect to the plurality of element type declarations of the markup language file.

          Moreover, element instances that are not changed in the  
25 markup language file are preferably not recorded in the delta file. Also, each change recorded in the delta file is preferably recorded at the lowest element level.

          The present invention additionally provides a method of generating a delta file containing data representative of one or more  
30 changes to a markup language file having application-defined tags, the delta file being created and changes to the markup language file being

recorded in the delta file where the changes are validated against substantially the same markup language structure as the markup language file to be changed.

In a further aspect the present invention provides software for  
5 recording and validating changes to a markup language file having application-defined tags, the software consisting of a set of instructions for creating a delta file and recording in the delta file changes to the contents of the markup language file where the recorded changes are validated against substantially the same markup language structure as the markup language  
10 file to be changed.

The present invention also provides software for generating a delta file containing data representative of one or more changes to a markup language file having application-defined tags, the software consisting of a set of instructions for creating the delta file and recording in  
15 the delta file changes to the content of the markup language file, the changes being being validated against substantially the same markup language structure as the markup language file to be changed.

The method of the present invention considers a markup language file as a structure of tagged data elements where the differences  
20 that are not significant, e.g. some occurrences of white space, can be ignored and where differences in the ordering of data can be ignored (unless, of course, they are significant). This makes the method particularly suited for use with more complex data files such as those used in CAD systems.

25 With the present invention the delta file, which contains a record of changes to be made or already made to a markup language file, can be viewed and processed as a markup language file in the normal manner. Also, the contents of the delta file follows the same nested markup structure as the structure of the markup language file that is  
30 changed. Although a DTD defining the structure of an XML file, for example, is not suitable for representing changes to files conforming to that

DTD, with the present invention it is possible to ensure strict compliance of recorded changes with the DTD of the markup language file, particularly in the case of more complex data files, by generating a delta file that has a DTD which is only slightly modified with respect to the DTD of the original XML file. As the DTD of the delta file constrains the contents of the delta file and as the delta DTD is based upon and is only a slightly modified version of the original XML file DTD, changes recorded in the delta file, that are potential amendments to the original XML file, are constrained to conform substantially to the DTD of the original XML file. In this way the delta file generated with the present invention is able to prevent, as far as possible, invalid changes to the XML file being represented in the delta file. For example, where the original DTD specified a choice between two elements, and exactly one was required to be present in the file, the delta DTD would similarly require exactly one of these two elements or an exchange sequence where each item in the exchange is required to be one of the two original elements.

It will of course be immediately apparent that a file conforming to the DTD of the delta file may represent the differences between two existing markup language files or may represent a set of changes that are to be applied to an existing markup language file or differences that were previously applied to an existing markup language file. Furthermore, the contents of the delta file can be chosen to be bi-directional in which case the same delta file can be used to generate either markup file from the other or can be constrained so that the delta file can only be applied to one of the two markup files to generate the other.

The approach of basing a delta DTD on the DTD of the original markup language file has a number of advantages. These and other advantages of this invention include but are not limited to the following:

- the validation process ensures that the changes represented in the delta file are well-formed and, optionally, strictly compliant with respect to a

DTD thus providing a robust change-control representation which is tightly coupled to the original DTD;

- processors designed to process files conforming to the original DTD can easily be modified to work on files conforming to the delta DTD;
- 5 • the delta file can be understood easily by a user familiar with the structure of the original files because the look and feel of the original DTD is maintained in the delta DTD;
- the size of the delta file is kept as small as possible;
- the delta file can be used as a basis for one or both of 'update' (convert  
10 the old version to the new) and 'undo' (convert the new version to the old) operations;
- multiple delta files can be incorporated into a single markup language document/file thus creating a revision history file;
- the delta file may contain the original data in addition to the changes  
15 and thus be the sole basis for generating one or other of the original files. Such a delta file is also useful in mergers of the two files where the operations to generate these can be defined as generic operations/transformations, for example using an XSL (XML Stylesheet Language) stylesheet that may be applied to any delta file
- 20 • data that is inherently order-independent can be handled;
- data that has 'keys' can be handled.

With the present invention, for example, a method is provided for making minimal modifications to any element type declaration so that it is capable of representing the differences between any two XML element  
25 instances which are both valid with respect to the original element type declaration. Such a process will work for any arbitrarily complex XML element type declaration or indeed for other markup language structures that are based on regular expressions with deterministic content models. The modifications are made in such a way that element instances which  
30 conform to the original element type declaration will also conform to the modified element type declaration.

The method can be applied to all element type declarations that define a particular document type and thus a delta format for that document type can be defined. Also, optimisations can be obtained by recursive application of the process from the top-level element type  
5 declaration down through the elements which it references to include only those changes that, for reasons related to the precise structure of the element type declarations, can occur in the particular document type.

In addition the method takes into account two common requirements of data structures which are not represented within a standard  
- 10 XML element type declaration, namely the ability to declare repeated items as being order-independent and to declare keys by which repeated element instances can be identified.

XML attributes are considered to be part of the data, and this means that a change to the value of an attribute needs to be recorded.  
15 Changes to the order of attributes are not significant.

The present invention is suited to applications involving the storage and communication of markup language data files particularly where the communication of data is over a network such as the internet. Furthermore, the present invention is suited to the merger of related files  
20 and the extraction of files from a merged document. Indeed, in a preferred embodiment the present invention provides a method and software for generating a delta file in which is stored a markup language file and changes to be made or that have been made to that file.

Embodiments of the present invention will now be described  
25 by way of example with reference to the accompanying drawings, in which:

Figure 1 illustrates the hardware requirements for a change recordal and validation software product;

Figure 2 is a flow diagram of operational steps performed by change recordal and validation software with respect to the identification  
30 and validation of changes between two XML files; and



Figure 3 is a flow diagram of operational steps performed by change recordal and validation software with respect to user input changes to an XML file.

As mentioned above the structure of an XML file is specified  
5 by a set of element type declarations in a DTD. Any amendment of the XML file is only valid if it is valid for its DTD i.e. the XML file before the amendment is valid for its DTD and the XML file after the amendment is also valid for its DTD. The change recordal and validation software utilises the DTD of an XML file that is to be changed or an XML file that has already  
10 been altered in the generation of a delta file that records changes to the XML file. XML attributes are considered to be part of the data and so a change to the value of an attribute also needs to be recorded. However, changes to the order of attributes are not significant and are not recorded as changes.

15 As illustrated in Figure 1, the change recordal and validation software requires the following resources: a processor 1 for carrying out the functional operations defined in the software; program storage 2; a data storage device 3 which for ease of reference may be sub-divided into first 4, second 5 and third 6 etc. storage regions; additional RAM 7 for temporary  
20 storage of data whilst the software is running; a comparator 8 for comparing the contents of XML files; and user input and output interfaces 9, 10. Most conveniently, the software is implemented on a conventional personal computer. However, the method of change recordal and validation of markup language files using application-defined tags described herein is  
25 not restricted to such an implementation.

### **Validation of Differences between Two XML Files**

A method of recording differences between two XML files and ensuring strict compliance of the differences with the DTDs of the XML files using the change recordal and validation software described herein is  
30 illustrated in Figure 2. The first and second XML files, that have differences between them, are either already stored in respective first and second

storage regions 4, 5 of the data storage device 3 or are input by means of the input interface 9 to the data storage device 3.

Firstly a delta file must be created to which changes between the two XML files can be written. To create the delta file the change  
5 recordal and validation software is enabled 11 and the processor reads 12 the document type definition (DTD) of the first XML file and creates a copy 13 of the DTD. The copy of the DTD is validated 14 using conventional techniques. In the event the DTD is found not to be valid the program ends 15 and an error message may be displayed by means of the output  
10 interface 10. Where the copy of the DTD is found to be valid, a modified DTD which will be the delta DTD is generated 16. To generate the delta DTD each element type declaration in the copy of the DTD is read 17 and analysed to identify whether the element type declaration requires amendment. Rules pertaining to modification of each of the element type  
15 declarations in the generation of the delta DTD are illustrated in Table 1 below. In turn each element type declaration is either retained in its original form in the delta DTD or is modified in accordance with the rules and the modified version is written to the delta DTD 18. This process is repeated 19 until all element type declarations of the original DTD have been read  
20 and retained unamended or amended and the complete delta DTD is saved 20 to the third storage region 6 of the data storage device 3. The delta DTD may be saved as a separate file or may be saved as part of the XML delta file in which the differences between the two XML files are to be recorded.

25 Now that the delta DTD has been generated, data which represents differences between the first and second XML files can be written to an XML file that is constrained to the delta DTD. The first and second XML files are read 21 from storage regions 4, 5 and using the RAM 7 and comparator 8 the contents of the two XML files are compared 22 to  
30 identify differences between them. Where a difference between the two XML files is identified 23, the relevant element type declaration is checked

to determine whether the difference is significant i.e. a semantic difference  
24. If the difference is not significant then the processor 1 continues  
checking for any more differences. If the difference is a semantic  
difference, using the delta DTD a valid representation of the difference is  
5 generated 25 and written 26 to the delta file. The processor 1 then checks  
for any further differences 23. If a valid representation of the difference  
cannot be generated, i.e. the difference is found not to be valid for the delta  
DTD, the program ends and an error message may be issued via the output  
interface 10. This process is repeated until all differences between the two  
-10 XML files have been analysed and, where appropriate, the differences  
recorded in the delta file. Once no further differences are identified the  
program ends 27.

All semantic differences between the two XML files are  
recorded in the delta file and these include changes to the textual/visual  
15 data in the XML document as well as formatting and other command  
language contained in the XML files. It will, of course be apparent that this  
method may be altered so that instead of ending where an invalid  
difference is identified, in addition to generating an error message that is  
output via the output interface 10, the program may continue, skipping the  
20 invalid difference which may be stored in a separate text file for accessing  
by a user once the program has ended.

#### **User Editing of an XML File**

The change recordal and validation software may also be  
used to record changes to a markup language file entered by a user via the  
25 input interface 9 or similarly changes made directly by an application  
program. Figure 3 illustrates the method steps for recording changes to an  
XML file, like reference numerals are used where appropriate. With this  
method a single markup language file such as an XML file is stored in the  
first storage region 4 of the data storage device 3. The software is started  
30 11 and, as before, initially the delta DTD is generated in the same manner  
as described above and is saved 20.

The XML file is read and a temporary copy made and stored 28 in the RAM 7. The temporary copy of the contents of the XML file is displayed 29 by means of the output interface 10 so that a user is able to input modifications to the XML file via the input interface 9. Each of the  
5 changes entered by the user is compared 30 to the temporary copy of the XML file and checked 31 to establish whether the modification is significant i.e. a semantic change. Where the modification is not significant the change is written 32 to the temporary version of the XML file in the RAM 7 but no record of the change is entered in the delta file and the processor 1  
10 checks 33 whether any further changes have been entered by the user.

Where the modification is identified as a semantic change, the processor checks 34 whether a valid difference representation of the change can be generated using the delta DTD. Where a valid difference representation is possible, the difference representation is written 35 to the delta file and the  
15 change written to the temporary copy of the XML file 32 with the processor checking 33 for any further modifications entered by the user. If the change proposed by the user is invalid, an error message is issued via the output interface to the user and the change is deleted from the delta file and from the temporary XML file in the RAM 7 and the program ends. Alternatively,  
20 whilst deleting the change from the delta file and the temporary file, the program may issue an error message via the output interface to the user.

Once the user has finished entering modifications to the XML file and all of the modifications have been found to be either not significant or valid semantic changes, the temporary version of the XML file in the  
25 RAM 7 is written over the original XML file in the first storage region 4. Of course, the modified version of the XML file may be stored separately from the original version of the XML file instead of overwriting the original XML version.

As the delta file contains all of the valid semantic  
30 modifications to the original XML file, an equivalent to the original XML file can be recreated from the modified XML file through reverse

implementation of the changes stored in the delta file. In this way the delta file can be used as an historical record of changes to an XML file.

Where a delta DTD is not generated, the content of the delta file is made to reflect the structure of the original files and to be well-formed.

5 Thus, in the cases of Figures 2 and 3 the method steps illustrated on the right-hand side of the figures are performed in the same way although the difference representation performed in box 25 and in box 34 will be based on a content model of ANY rather than a content model specified in a DTD.

#### **Generation of the Delta DTD**

10 In accordance with the methods described above each element type declaration governing the first XML file is examined and modified by the processor 1 to create the delta DTD. Below, in Table 1, are set out the core rules that are applied in defining the changes that must be made to an XML DTD in order to create a derived DTD that will govern the  
15 delta file. The DTD of the delta file and the DTD of the original markup language file are substantially identical in that the nested structure of the contents of the delta file and the original file will be substantially the same. However, certain slight modifications are necessary to enable the delta file to represent the changes to a markup language file. All of the element type  
20 declarations are either copied across to the delta DTD or amended and the element type declarations are processed to the lowest level of each content particle. The rules provide one embodiment of an implementation but this does not preclude others based for example on the generation of different element declarations rather than the use of attributes and/or the use of XML  
25 Namespaces (<http://www.w3.org/TR/REC-xml-names>) and/or the use of XML Schema (<http://www.w3.org/TR/xmlschema-1>) the contents of which are incorporated herein by reference. Table 2 illustrates how unique XML Schema features would be represented in a delta file. In other respects, the process of creating a delta XML Schema to validate differences between  
30 two files conforming to an XML Schema specification follows the same principles as that of creating a delta DTD except that the syntactic

representation of DTDs and XML Schemas of course differ.

These tables show how different types of content particle used within elements are modified to generate the delta DTD or XML Schema . In the case of Table 1 a content particle is part of an element  
5 type declaration and consists of element names, choice lists of content particles, or sequence lists of content particles.

### **New Elements**

As mentioned above, the delta DTD includes certain new elements. The names of these new elements are not significant but the  
10 purpose of each is important and needs to be known to the processing system. This could be achieved in a number of ways including the use of fixed names or fixed XML attribute values. In Table 3 below, the new elements are assumed to be named as in the table but this is not a requirement. The use of namespaces, especially when using XML  
15 Schema, provides a way of ensuring that the names of these elements and any additional attributes do not clash with any names in the original files, e.g. by applying a 'deltaxml:' prefix linked to a unique namespace.

Table 1

**Changes to Content Particles**

Content Particle Type	Description	Example(s)	Delta CP for Example	DTD Modifications
parentheses	Simple bracket containing a single item, i.e. element or content particle	(x) ((x, y))	(x) ((x, y))	No change. Content particles within the parentheses are processed according to their structure.
required element	Item is an element which is not ?*+	x y	x y	No change.
optional element	Item is an element which is optional, ?	x? y?	x? y?	No change.
repeated optional element	Item is an element which is *	x* y*	x* y*	No change.
repeated required element	Item is an element which is +	x+ y+	(x+   repeatedItemsUnchanged)	The repeated required element is incorporated into a new choice content particle that also allows repeatedItemsUnchanged.
sequence	Content particle is a sequence with no ?*+	(x, y) (x, (y, z)?, a)	(x, y) (x, (y, z)?, a)	No change. The content particle for each sequential item is processed according to its structure.
optional sequence	Content particle is a sequence with ?	(x, y)? (x, (y, z)?, a)?	(x, y)? (x, (y, z)?, a)?	No change. The content particle for each sequential item is processed according to its structure.
repeated optional sequence	Content particle is a sequence with *	(x, y)* (x, (y, z)?, a)*	(x, y)* (x, (y, z)?, a)*	No change. The content particle for each sequential item is processed according to its structure.
repeated required sequence	Content particle is a sequence with +	(x, y)+ (x, (y, z)?, a)+	((x, y)+   repeatedItemsUnchanged) ((x, (y, z)?, a)+   repeatedItemsUnchanged)	The content particle for each sequential item is processed according to its structure and the resultant repeated required sequence is incorporated into a new choice content particle that also allows repeatedItemsUnchanged.

required choice	Content particle is a choice with no ?*,+	(x   y   z)	((x   y   z   (exchangeStartSingle, (x   y   z), exchangeMiddle, (x   y   z), exchangeEnd))	A new exchange sequence is formed where the content particle between the exchangeStartSingle and exchangeMiddle is the original content particle, and the content particle between the exchangeMiddle and exchangeEnd is the original content particle. Each of the items in the original choice is processed according to its structure, and a new required choice content particle formed from these items and the new exchange sequence.
optional choice	Content particle is a choice with ?	(x   y   z)?	((x   y   z   (exchangeStartSingle, (x   y   z), exchangeMiddle, (x   y   z), exchangeEnd)))?	The processing is as for a required choice except that the result is an optional choice rather than a required choice.
repeated optional choice, ordered	Content particle is a choice with *	(x   y   z)*	((x   y   z   (exchangeStart, (x   y   z), exchangeMiddle, (x   y   z), exchangeEnd))*	The processing is as for a required choice except that the result is an repeated optional choice rather than a required choice.
repeated optional choice, unordered	Content particle is a choice with *	(x   y   z)*	(x   y   z )*	<b>No change.</b> The content particle for each choice item is processed according to its structure.
repeated required choice, ordered	Content particle is a choice with +	(x   y   z)+	((x   y   z   (exchangeStart, (x   y   z), exchangeMiddle, (x   y   z), exchangeEnd))+   repeatedItemsUnchanged)	The processing is as for a required choice except that the result is an repeated optional choice rather than a required choice, and this is then incorporated into a new choice content particle which also allows repeatedItemsUnchanged.
repeated required choice, unordered	Content particle is a choice with +	(x   y   z)+	((x   y   z )+   repeatedItemsUnchanged)	The content particle for each choice item is processed according to its structure, and the result is a repeated required structure which is incorporated into a new choice content particle which also allows repeatedItemsUnchanged.



Table 2

XML Schema feature	Delta Schema Modifications
A combination type, "all", can be applied to a set of elements to indicate that each may occur at most once but in any order.	The delta Schema for an 'all' is the same as the original.
MIXED content can be controlled more precisely.	Handled as MIXED in deltaDTD, ignoring the structure of the elements. For improved validation a more complex delta Schema could be generated.
Local elements can be defined.	These elements can be processed into delta elements in the same way as non-local elements but the resultant delta element remains as a local element definition.
Restrictions, extensions, groups and other structural components	These need to be expanded out before a delta process is applied.
PCDATA and attribute types	Handled as in delta DTD but more sophisticated comparison functions can be written to detect the case where items are the same, e.g. "6 June 2001" and "6/6/2001". For improved validation, additional elements can be defined to handle changes to structured PCDATA in a structured way.
namespaces	Namespaces are handled in XML Schema but not in DTDs. Generation of a delta schema can be applied to each schema file which represents a target namespace.
substitution groups	These can be resolved to choices and treated as normal choices.
Element occurrences can be controlled by minOccurs/maxOccurs with numeric values other than 0 or 1 (DTD only allows 0, 1 and 'unbounded' for maximum)	Can be treated as '*' or as '+' as appropriate. For improved validation, occurrence limits could be set in the delta schema.

Table 3

Element type name	Content	Purpose
exchangeStart	EMPTY	Introduces the start of an exchange sequence. In some complex content particles, it is necessary to have more than one element of this type declared. A suffix may be added to distinguish cases where the exchange sequence cannot be repeated, e.g. exchangeStartSingle
exchangeMiddle	EMPTY	Denotes the mid-point of an exchange sequence
exchangeEnd	EMPTY	Denotes the end-point of an exchange sequence
PCDATAmodify	(PCDATAold, PCDATAnew)	Records a change to PCDATA
PCDATAold	#PCDATA	Records the original PCDATA content
PCDATAnew	#PCDATA	Records the new PCDATA content
MIXEDmodify	(MIXEDold, MIXEDnew)	Records a change to MIXED content
MIXEDold	ANY	Records the original MIXED content
MIXEDnew	ANY	Records the new MIXED content
repeatedItemsUnchanged	EMPTY	Records that for a content particle that is repeated at least once (denoted by '+' in the DTD), all items were unchanged. In some complex content particles, it is necessary to have more than one element of this type declared.

5

In addition to instances of the above elements, each element instance in a delta file is present for one of a number of different reasons. These are enumerated in Table 4 and are known as difference types:

Table 4

Difference type	Meaning	Constraints
originalData	No differences – this is part of the original data, either in the 'old' file or the 'new' file.	Applies to all element instances which are contained in any element instance with "unchanged", "add" or "delete".
unchanged	Specifies that the element instance is unchanged.	No nested element instances or attributes are needed within the element instance.
add	Specifies that this element instance is to be added (present in the 'new' file but not in the 'old').	All element instances between an exchangeMiddle and exchangeEnd will have this value.
delete	Specifies that this element instance is to be deleted (present in the 'old' file but not in the 'new').	All element instances between an exchangeStart and exchangeMiddle will have this value.
modify	Specifies that this element instance is modified	Specifies that this element instance is modified, so all element instances directly within this will either correspond to elements in Table 2 or will have values of "unchanged", "add", "delete" or "modify".

The top-level element instance in a document could have a  
5 difference type of "unchanged" or "modify". Element instances below the  
top-level element instance will follow the constraints given in the tables.  
Difference types do not apply to instances of elements in Table 3.

The difference type of each element in the delta file also  
needs to be represented in some way, e.g. using an attribute value. In the  
10 examples below, an attribute with the name "delta" is used and has values  
according to the table above, except that originalData is not specified but is  
taken as the default value.

The following are examples of valid changes that may be  
made to the element type declarations of a DTD with reference to their  
15 content. Elements whose content is #PCDATA will have their content  
changed from:

(#PCDATA) or (#PCDATA)\*

to (#PCDATA | PCDATAmodify)\*

For example, the element type declaration:

20 <!ELEMENT comment (#PCDATA)>

would become:

<!ELEMENT comment (#PCDATA | PCDATAmodify)\* >

The representation of the differences between:

<comment>This is a comment</comment>

and:

5                   <comment>This is another comment</comment>

would be represented in the delta file as:

<comment delta="modify"><PCDATAmodify>

<PCDATAold>This is a comment</PCDATAold>

<PCDATAnew>This is another comment</PCDATAnew>

10                   </PCDATAmodify></comment>

It is also possible to use the structure for a finer-grained change, for example:

<comment delta="modify">This is <PCDATAmodify>

<PCDATAold>a </PCDATAold>

15                   <PCDATAnew>another </PCDATAnew>

</PCDATAmodify>comment</comment>

The above examples would create a bi-directional delta file.

Where the delta file is only required to detail changes from the 'old'

20   document to the 'new', the same change could be represented in the delta file as:

<comment delta="modify"><PCDATAmodify>

<PCDATAold/>

<PCDATAnew>This is another comment</PCDATAnew>

25                   </PCDATAmodify></comment>

As the structure of the delta file follows the structure of the 'old' file and so is ordered and well-formed, the relevant comment in the 'old' file that is to be modified is automatically identifiable. Of course, a similar directional delta file could be generated that contains only the information necessary to convert a 'new' file back into the 'old' file. This

30

enables the delta file to be even smaller in size than one that is bi-directional. This optimisation cannot always be applied, e.g. when some of the elements in the file are treated as unordered and do not have keys.

MIXED content elements can be treated in a similar manner to #PCDATA. Note that the items within the MIXEDmodify element may have elements as well as PCDATA in them. For example:

```
<!ELEMENT mixedContentExample (#PCDATA | a | b)*>
```

would become:

```
<!ELEMENT mixedContentExample (#PCDATA |a|b|MIXEDmodify |  
PCDATAmodify)*>
```

MIXED content is generally considered to be ordered.

Content of ANY is similar to MIXED and so is handled as follows.

```
<!ELEMENT anyContent ANY>
```

would become:

```
<!ELEMENT anyContent ANY>
```

but within this, MIXEDmodify or PCDATAmodify would be used to denote changes. As for MIXED content, ANY is also generally considered to be ordered.

20

A set of three element types denotes exchanges of element instances within a document. An exchange is the replacement of zero or more element instances, which conform to a content particle in the element declaration, with another list of zero or more element instances conforming to the same content particle definition. An exchange sequence is represented as:

25

```
<exchangeStart>xx<exchangeMiddle>yy<exchangeEnd>
```

where 'xx' is the data for zero or more element instances from the first file, and 'yy' the data for the same content particle in the second file. Each element within xx will have a delta attribute set to "delete" and each element within yy will have a delta attribute set to "add". It should be noted that in some complex nesting of content particles, it is necessary to have

30

different exchangeStart elements in order to keep the element content model deterministic. In these cases new exchangeStart elements could, for example, be generated using a '1', '2' etc. suffix.

Where optional items, denoted in an XML DTD using '\*' and  
5 '?' , are equal in the first and second XML files there will be nothing in the delta file.

However, where an item is required in the DTD of the original XML file, there will always be something present in the delta file. For a single required item, where there are no differences between the first and  
10 second XML files, the element that was present in both files will also appear in the delta file and it will have a delta attribute with value "unchanged", and no content. This means that any element that could be unchanged in the delta file must have an element declaration that allows it to be empty or is amended to allow it to be empty.

15 In the case of a repeated required item, denoted by '+' in an XML DTD, if all items are equal in both files, there will be a single element in the delta file:

`<repeatedItemsUnchanged/>`

It should be noted that in some complex nesting of content  
20 particles, it is necessary to have different repeatedItemsUnchanged elements in order to keep the element content model deterministic. In these cases new repeatedItemsUnchanged elements could, for example, be generated using a '1', '2' etc. suffix.

The order of elements within an XML document is normally  
25 significant. However, in many cases the order of data elements may not be significant and this method allows for such a situation to be specified so that it can be handled in the generation of the delta DTD and in the comparisons. The precise way in which this is specified is not important. Any repeating content particle, including individual elements, may be  
30 specified to be unordered. In this case the occurrences of this content

particle in the document may occur in any order. However, this does not imply that the content particles within it are unordered.

The effect on the generation of the delta version of a content particle is that the exchange form is not required, as noted in Table 1.

5 XML has no concept of keys to the items within a document although this is a common feature of many data structures. Therefore this method allows the possibility that an item may have a key. A key is defined as some aspect of an element that is unique within the element that contains it. Two element instances that correspond with the same element  
10 type declaration, are in the same content particle occurrence in two documents, and have equal keys are considered to represent the same item, i.e. can be compared.

The effect of this on the delta document is that the constituent parts of the key need to be present in the delta XML file so that navigation  
15 can be repeated through the structure. There is no direct effect on the delta version of a content particle due to the presence of keyed items within it. There is no restriction on what constitutes a key: this could include attributes or elements and these may be required or optional.

Attributes are treated in the same manner as data in that any  
20 change will be recorded in the delta file.

Attributes of type ID and IDREF in the original DTD are changed to be of type NMTOKEN in the delta file. Similarly, attributes of type IDREFS are changed to be of type NMTOKENS. The reason for this is that the delta file may contain two ID attributes with the same value,  
25 because they come from different files. Similarly, an IDREF attribute may reference an ID that is not present in the delta file, so it cannot remain of type IDREF.

Attributes that were #REQUIRED are changed to be #IMPLIED. The reason for this is that unchanged attributes are not present  
30 in the delta file unless they are part of a key.

Any changes to attributes need to be recorded. This could be done in any way, e.g. by using two new attributes which contain in an encoded form both the names and the values of the attributes. With this implementation:

- 5 • one attribute contains the details of all attributes in the original file which are not present or have changed in the new file,
  - the other attribute contains the details of all attributes in the new file that were absent or different in the original file, and
  - unchanged attributes do not normally need to appear in the
- 10 delta file.

A characteristic of the delta file is that the changes specified can be applied successfully to files other than the original files that were used to generate the delta file. This might enable, for example, changes made to a file A to produce B to be applied to a new version of A, denoted

15 as A' to produce B'. This works predictably when certain conditions are met regarding the relationship between A and A'. For example, the file  $\Delta$  which represents the differences between two files A and B can be applied to another file A' which is some derivative of A such that:

- all items in A that are modified in  $\Delta$  are present in A'
- 20 • the position in the XML tree of items modified in  $\Delta$  are the same in A' as in A, where position may be related to order or keys as appropriate.

The following are some examples of element declarations and the way that they will be modified in the delta DTD. Instances of the

25 elements are also given to show how the differences are represented. In the case of a single required element within another element an example of an element declaration is:

```
<!ELEMENT ex1 (x)>
<!ELEMENT x (#PCDATA)>
```

30 The resultant delta element declaration is:

```
<!ELEMENT ex1
```



```
(x)?>
<!ATTLIST ex1
  delta (modify | unchanged | originalData) #IMPLIED
  new-attributes CDATA #IMPLIED
5  old-attributes CDATA #IMPLIED
  <!ELEMENT x
    (#PCDATA | PCDATAmodify)*>
  <!ATTLIST x
    delta (modify | unchanged | originalData) #IMPLIED
10  new-attributes CDATA #IMPLIED
  old-attributes CDATA #IMPLIED
```

In the above, as x is required it cannot be added or deleted and so the delta attribute does not allow these values. The delta element  
15 declaration makes (x) optional so that in the case where ex1 is unchanged, no data need be included within it.

For example:

Input file 1: <ex1><x>a b c</x></ex1>

20 Input file 2: <ex1><x>x y z</X></ex1>

Delta file element:

```
<ex1 delta="modify">
  <x delta="modify"><PCDATAmodify>
  <PCDATAold>a b c</PCDATAold><PCDATAnew>x y z</PCDATAnew>
25  </x></ex1>
```

For a repeated optional element that is within another element and where the order is not significant, the information may be represented using a FIXED attribute.

For an element declaration:

```
30  <!ELEMENT ex2 (x*)>
    <!ATTLIST ex2 not-ordered CDATA #FIXED "x">
    <!ELEMENT x (#PCDATA)>
```

The resultant delta element declaration is:

```
<!ELEMENT ex2
(x*)>
<!ATTLIST ex2
5  delta (modify | unchanged | originalData) #IMPLIED
   new-attributes CDATA #IMPLIED
   not-ordered CDATA #FIXED "x"
   old-attributes CDATA #IMPLIED>
<!ELEMENT x
10  (#PCDATA )>
   <!ATTLIST x
      delta (add | delete | originalData) #IMPLIED
      new-attributes CDATA #IMPLIED
      old-attributes CDATA #IMPLIED>
```

15 For example:

Input file 1: <ex2><x>a b c</x><x>x y z</x></ex2>

Input file 2: <ex2><x>!x y z!</x><x>a b c</x><x>c d e</x></ex2>

```
Delta file element:  <ex2 delta="modify">
                    <x delta="delete">x y z</x>
20                  <x delta="add">!x y z!</x>
                    <x delta="add">c d e</x>
```

There are some differences here as compared to the ordered case. As the instances of x are not in order, there is no way to work out which ones correspond, it is only possible to say which ones are present in both files and which are not. Therefore, it is not possible to 'modify' any particular instance, it is only possible to identify instances that have been added and ones that have been deleted. Occurrences that have not changed will not appear in the delta file, so there is no possibility of having an instance of x with a delta attribute of 'unchanged'.

30 The following example illustrates how it is possible to optimise the delta DTD, that is to avoid allowing for changes in elements that cannot be changed because of the way that they are used.

Thus the element declaration might be:

```
<!ELEMENT ex3 (x+)>
<!ATTLIST ex3 not-ordered CDATA #FIXED "x">
<!ELEMENT x ((y | z)+)>
5 <!ELEMENT y (#PCDATA)>
  <!ELEMENT z (#PCDATA)>
```

The element instances of x within ex3 are not ordered.

Therefore any instance that is unchanged will not appear in the delta file,  
and as it is not possible to identify which instances have changed, it is also  
10 not possible to modify an instance of x. By implication therefore, it is not  
possible to have a modification of y or z.

The resultant delta element declaration is thus:

```
<!ELEMENT ex3
((x+ | repeatedItemsUnchanged))?>
15 <!ATTLIST ex3
  delta (modify | unchanged | originalDATA) #IMPLIED
  new-attributes CDATA #IMPLIED
  not-ordered CDATA #FIXED "x"
  old-attributes CDATA #IMPLIED
20 <!ELEMENT repeatedItemsUnchanged EMPTY>
  <!ATTLIST repeatedItemsUnchanged
    deltaElement CDATA #FIXED "repeatedItemsUnchanged">
  <!ELEMENT x
    (y | z)+>
25 <!ATTLIST x
  delta (add | delete | originalData) #IMPLIED
  new-attributes CDATA #IMPLIED
  old-attributes CDATA #IMPLIED>
  <!ELEMENT y
30 (#PCDATA)>
  <!ELEMENT z
    (#PCDATA)>
```

In the delta DTD, the only possible delta attribute values for x

are add and delete. As x cannot be modified, its structure has not been changed and items below it do not need to have a delta attribute as they will always be part of the original data.

Input file 1:

```
5      <ex3>
      <x><y>a b c</y></x>
      <x><z>a b c</z></x>
      <x><y>c d</y></x>
      </ex3>
```

10 Input file 2:

```
      <ex3>
      <x><y>c d</y></x>
      <x><z>c b a</z></x>
      <x><y>a b</y></x>
15     </ex3>
```

Delta file element:

```
      <ex3 delta="modify">
      <x delta="delete"><y>a b c</y></x>
      <x delta="delete"><z>a b c</z></x>
20     <x delta="add"><z>c b a</z></x>
      <x delta="add"><y>a b</y></x>
      </ex3>
```

25 In this example, the instance `<x><y>c d</y></x>` appears in both files and so is not present in the delta file. The other two items are different and are therefore deleted and new items added.

In the examples given above the delta file contains only data on changes to a markup language file or between two markup language files. It is also possible for the delta file to contain all of the original data from a markup language file as well as the changes to be made to the file. Such a combination of the original data with the changes is referred to as a 'full' delta file and where a delta DTD has been generated the full delta file can be validated against this delta DTD. Full delta files are particularly

30

useful where mergers between files are required but have the disadvantage of being larger in size than standard delta files recording only changes. The following illustrates the differences between a standard delta file and a full delta file.

5

For:

```
<!ELEMENT ex4 (x*)>
<!ELEMENT x (#PCDATA)>
```

10 with a corresponding delta DTD of:

```
<!ELEMENT ex4 (x*)>
<!ELEMENT x (#PCDATA | PCDATAmody)*>
```

15 Input file 1: <ex4><x>first</x><x>second</x><x>third</x></ex4>

Input file 2: <ex4><x>first!</x><x>second</x><x>third</x></ex4>

The standard delta file in which unchanged data is omitted is:

20

```
<ex4 delta="modify">
  <x delta="modify"><PCDATAmody>
    <PCDATAold>first</PCDATAold>
    <PCDATAnew>first!</PCDATAnew></PCDATAmody>
  </x>
  <x delta="unchanged"/>
  <x delta="unchanged"/>
</ex4>
```

25

30 The full delta file in which all the data contents of elements are retained is:

```
<ex4 delta="modify">
  <x delta="modify"><PCDATAmodify>
    <PCDATAold>first</PCDATAold>
    <PCDATAnew>first!</PCDATAnew></PCDATAmodify>
5  </x>
  <x delta="unchanged">second</x>
  <x delta="unchanged">third</x>
</ex4>
```

10           The result of a comparison between two files can be represented as a full delta as described above containing both changes and the original data. The process for extracting from the full delta file either of the original files is straightforward and in effect the full delta is a merge of the two original files represented in a highly structured way such that the  
15   origin of any changed data and conflicts between the two files can be determined.

          It is possible to use a full delta file FD that represents the difference between two files A and B and where the FD is valid with respect to a delta DTD as a basis for merging the two files A and B such that the resultant  
20   merged file M is valid with respect to the original DTD. The merge can be achieved by including in the merged file M all data with delta values equal to "add" and "delete" and selecting either the data from A or the data from B in cases where an exchange is specified unless the exchanged content  
25   particles are permitted by the DTD to be repeated in which case both may be included. Typically user interaction is required to achieve a specific desired result unless it is always the case that either the old or new data is to take precedence for inclusion in the merged file. However, building the merge process on top of a valid full delta file simplifies the merge software considerably.

30           In the case where both A and B are themselves modifications of some common base file X, it is possible to use the delta process for a

three-way merge as follows. A delta file is generated from the difference between X and B and this delta file is then applied to A. Note again that some user interaction is needed to produce specific results where there are conflicts between the changes made in A and B.

5       To minimise user interaction the merger can be made fully automatic by distinguishing between an exchangeStart/Middle/End sequence in which a content particle may be repeated and such a sequence in which a content particle is allowed only once. For an automated merger if the content particle is allowed only once one or other must be included in the output  
10 file, otherwise both may be included. This is simply achieved by adopting a different element name for these two cases. With this change a generic script can be written to generate a merged file directly from a full delta file.

With the method and software for performing the method described above, changes of a markup language file, e.g. an XML file, that  
15 utilises application-defined tags may be defined in a separate file, the delta file, that is also an XML file and which conforms to a delta DTD derived from the DTD of the original XML file where a delta DTD is employed. In this way the constraints of the original XML file similarly apply to the delta file. As a result of this the delta file is able to validate the changes by  
20 ensuring the data recording the changes is well-formed or ideally strictly complies with the delta DTD, when a DTD is employed, thus providing a robust change-control representation which is tightly coupled to the original XML file. Moreover, users familiar with the structure of the original XML file will easily familiarise themselves with the delta file. Furthermore, by  
25 utilising the structure of the original XML file, data on the changes to the XML file can be compressed so that the delta file is made as small as possible. In particular, those element instances for which no change is recorded are noted as unchanged without compromising the structure of the original markup language file. To the extent that the delta file is  
30 concerned with semantic modifications only, the method and software described above analyses and generates a delta file on the basis of the

structure of the XML file as defined by the DTD rather than treating each line of the XML file as a line of text.

It will, of course, be apparent that the validation of changes to the markup language file is inherent from the structure assigned to the delta file. There is therefore no need for a separate validation step to be taken to ensure the changes to the markup language file are valid. The creation of a delta file, in accordance with the method described above, however unintentionally, provides validation of the changes proposed to a markup language file and recorded in the delta file. Of course, the validation of the changes only applies to potential amendments to the markup file. For example, where a change such as 'delete chapter 4' is recorded in the delta file but there is no chapter 4 in the original markup language file, this is not a potential amendment. However, if the original markup language file contained a chapter 4 that was a 'required' item then a change recorded in the delta file to delete chapter 4 may be found not to comply with the delta DTD i.e. the entry would not be a valid entry.

An application of the method and software for generating delta files is to permit multiple versions of a markup language file to be contained in a single markup file in such a way that only minimal data is held in the file and any version can be re-generated, as required. An example application for this is when a DTD is updated but a publisher wishes to publish data that conforms not only to the latest DTD but to previous versions also. The previous versions are all, typically, small variations of the latest version which makes it sensible to store them as delta files.

Consider a case of several versions of a data file V1, V2, V3, V4 where V4 is the latest version. Storing all of these in a single file would enable any to be retrieved but it would be too large a file. So, with the present invention, the following can be stored instead: delta(V4-V1), delta(V4-V2), delta(V4-V3) and V4. This would be a much smaller file and any version could be retrieved simply by applying the appropriate delta. For example, to generate V2 apply re-combination to (V4+delta(V4-V2)) to



get V2 again. It will be apparent, that with V4 included the delta files need be only unidirectional providing a further reduction in file size.

This approach works for any changes between versions other than for changes to the root element as there is no delta DTD involved in this.

5        Similarly, the possibility of storing delta files which enable different versions of data to be recreated is of direct application to database storage. Databases are now being revised to store XML directly and the present invention provides a way of storing XML in databases in a way that uses as little storage as possible. A delta file without a DTD would provide storage  
10        for any versions of XML data which in some cases may be the preferred option. A delta file with a DTD would provide more accurate and smaller delta files but the original files would, of course, have to conform to the same DTD.

      The present invention is also of relevance to the communication of  
15        data and in particular to the communication of data over networks. As delta files may be made considerably smaller than the actual data files that are being changed, the communication of delta files rather than the complete files over a network would be both more reliable and significantly faster. This is particularly important in the arena of the services offered by ASPs  
20        (Application Service Providers) over the web. An ASP may offer a service for performing a complex analysis of some engineering data. Very often the original data and the results are represented as XML data and in some cases these could be very large data files. Where the process is interactive, for example send data, get results, modify data, send new data,  
25        get new results etc., the present invention can greatly reduce the bandwidth required, improve the speed with which this data can be communicated between the user and the ASP which would have a consequential increase in the reliability in the data transfers. Moreover, the use of delta files in the transfer of information over a public network provides added security as an  
30        unauthorised recipient is not able to extract useable data from the delta file alone.

Delta XML middleware can be inserted in the pipeline between the user and the ASP so that only delta files need to be sent over the network. Of course, the full data file must be sent the first time but thereafter only delta files would be required. Furthermore, these are circumstances where  
5 a choice may be made between the adoption of bi-directional delta files or unidirectional delta files.

As can be seen there are applications of the present invention in the internet environment. Indeed, delta files may be used in the accessing and downloading of web pages over the internet. This is particularly of use with  
10 static web pages where part of the pages such as the header remains the same. For example, instead of a hyperlink on a web page identifying a new page to access, the hyperlink identifies the relevant delta file to be accessed which communicates only the changes to be made to the existing web page to generate the new page selected. This is particularly of  
15 significance with the increasing use of WAP (Wireless Application Protocol) and WML (Wireless Markup Language) and related telephony technology for accessing data from the internet etc.

It will, of course, be apparent from the examples given above that the present invention has many applications in the fields of data  
20 management and communication of which the above are only a cross-section and are non-limiting. Many other uses of the present invention are envisaged as are steps, other to those described above, that may be employed to generate delta files in accordance with the present invention.

## CLAIMS

1. A method of recordal of changes to a markup language file having application-defined tags wherein a delta file is created and changes  
5 to the contents of the markup language file are recorded in the delta file where the changes are validated against substantially the same markup language structure as the markup language file to be changed.
2. A method as claimed in claim 1, wherein the delta file is  
10 - created using the same markup language as the markup language file to be changed.
3. A method as claimed in either of claims 1 or 2, wherein the  
15 delta file is assigned a plurality of element declarations determined from the element declarations of the markup language file, the differences between one or more of the element declarations of the delta file and the element declarations of the markup language file being restricted to permissible differences with respect to the element declarations of the markup language file.  
20
4. A method as claimed in any one of the preceding claims, wherein element instances that are not changed in the markup language file are not recorded in the delta file.
- 25 5. A method as claimed in any one of the preceding claims, wherein each change recorded in the delta file is recorded at the lowest element level.
6. A method of generating a delta file containing data  
30 representative of one or more changes to a markup language file having application-defined tags, the delta file being created and changes to the

contents of the markup language file being recorded in a delta file where the changes are validated against substantially the same markup language structure as the markup language file to be changed.

5     7.             Software for recording changes to a markup language file having application-defined tags, the software consisting of a set of instructions for creating a delta file and recording in the delta file changes to the contents of the markup language file where the recorded changes are validated against substantially the same markup language structure as the  
10 --markup language file to be changed.

8.             Software as claimed in claim 7, wherein the delta file is created using the same markup language as the markup language file to be changed.

15

9.             Software as claimed in claims 7 or 8, wherein the set of instructions include the step of assigning to the delta file a plurality of element declarations determined from the element declarations of the markup language file the differences between one or more of the element  
20 declarations of the delta file and the element declarations of the markup language file being restricted to permissible amendments with respect to the element declarations of the markup language file.

10.            Software as claimed in any one of claims 7 to 9, including the  
25 additional instruction that element instances that are not changed in the markup language file are not recorded in the delta file.

11.            Software as claimed in any one of claims 7 to 10, wherein  
30 each change recorded in the delta file is recorded at the lowest element level.

12. Software as claimed in any one of claims 7 to 11, wherein the set of instructions includes instructions for additionally recording at least one of the markup language file and the changed markup language file.

5 13. Software for generating a delta file containing data representative of one or more changes to a markup language file having application-defined tags, the software consisting of a set of instructions for creating the delta file and recording changes to the content of the markup language file the changes being validated against substantially the same  
10 ---markup language structure as the markup language file to be changed.

14. Data storage medium on which is stored software for generating a delta file containing data representative of one or more changes to a markup language file having application-defined tags, the  
15 software consisting of a set of instructions for creating the delta file and recording changes to the content of the markup language file the changes being validated against substantially the same markup language structure as the markup language file to be changed.

20 15. A method of amending a later version of a markup language file to an earlier version of a markup language file comprising reading a delta file containing data on changes made to the earlier version of the markup language file wherein the changes are valid with respect to the markup language structure of the earlier markup language file, and applying  
25 the changes recorded in the delta file to the later version of the markup language file.

16. A method of amending a second markup language file comprising reading a delta file containing data on changes made to a first  
30 markup language file wherein the changes are valid with respect to the markup language structure of the first markup language file and are

potential amendments with respect to the content of the second markup language file and applying the changes recorded in the delta file to the second markup language file.

Figure 1

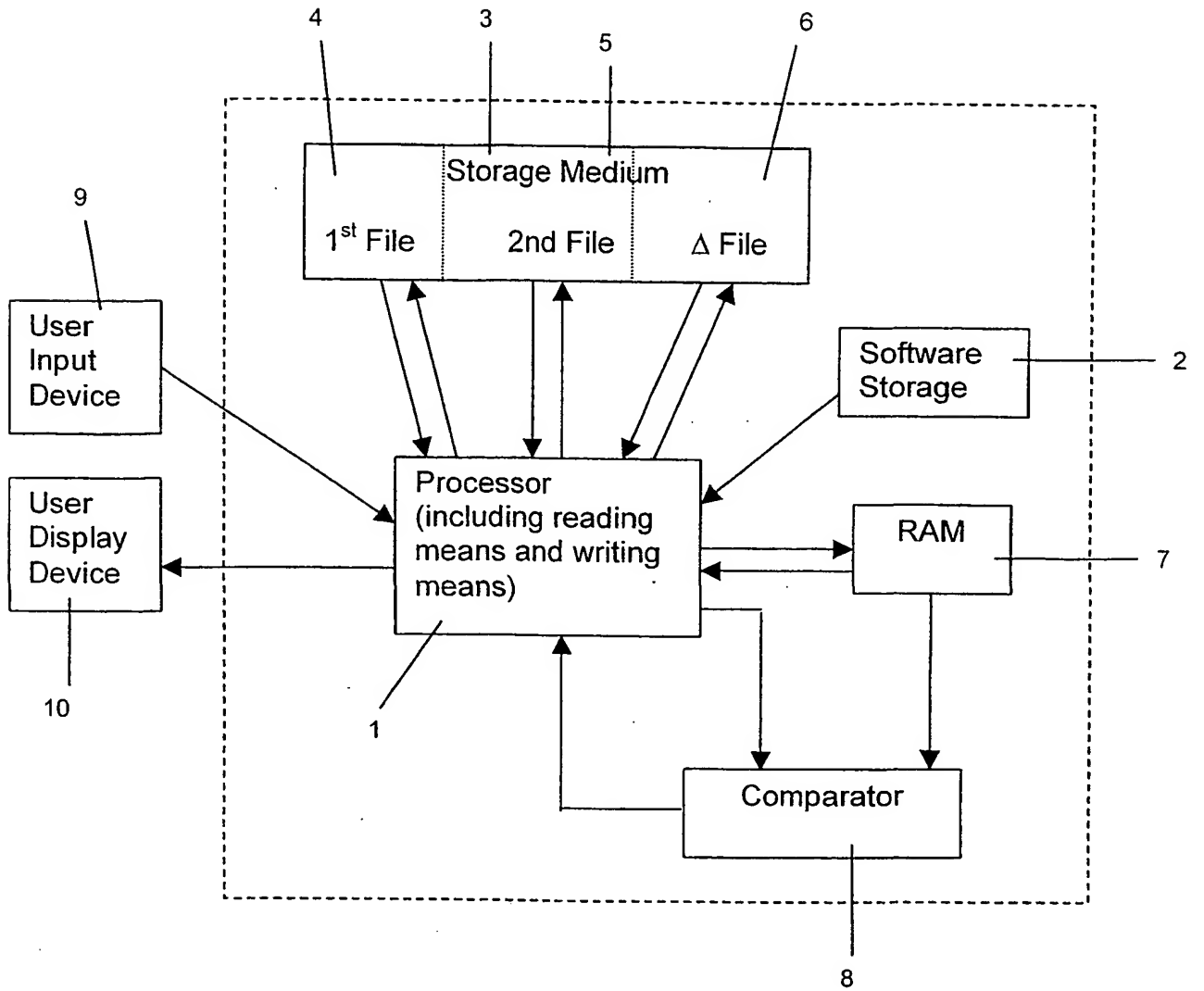


Figure 2

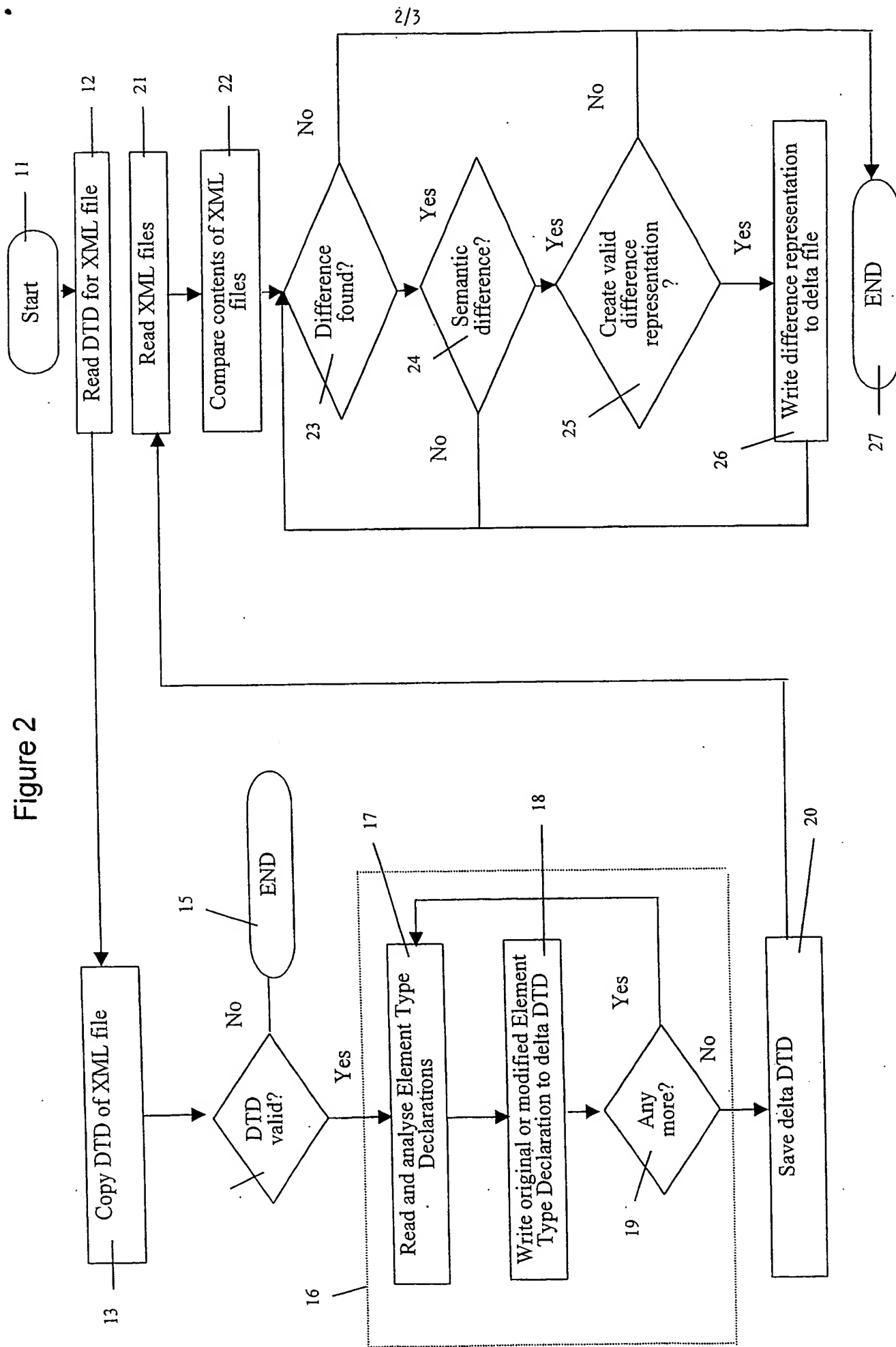
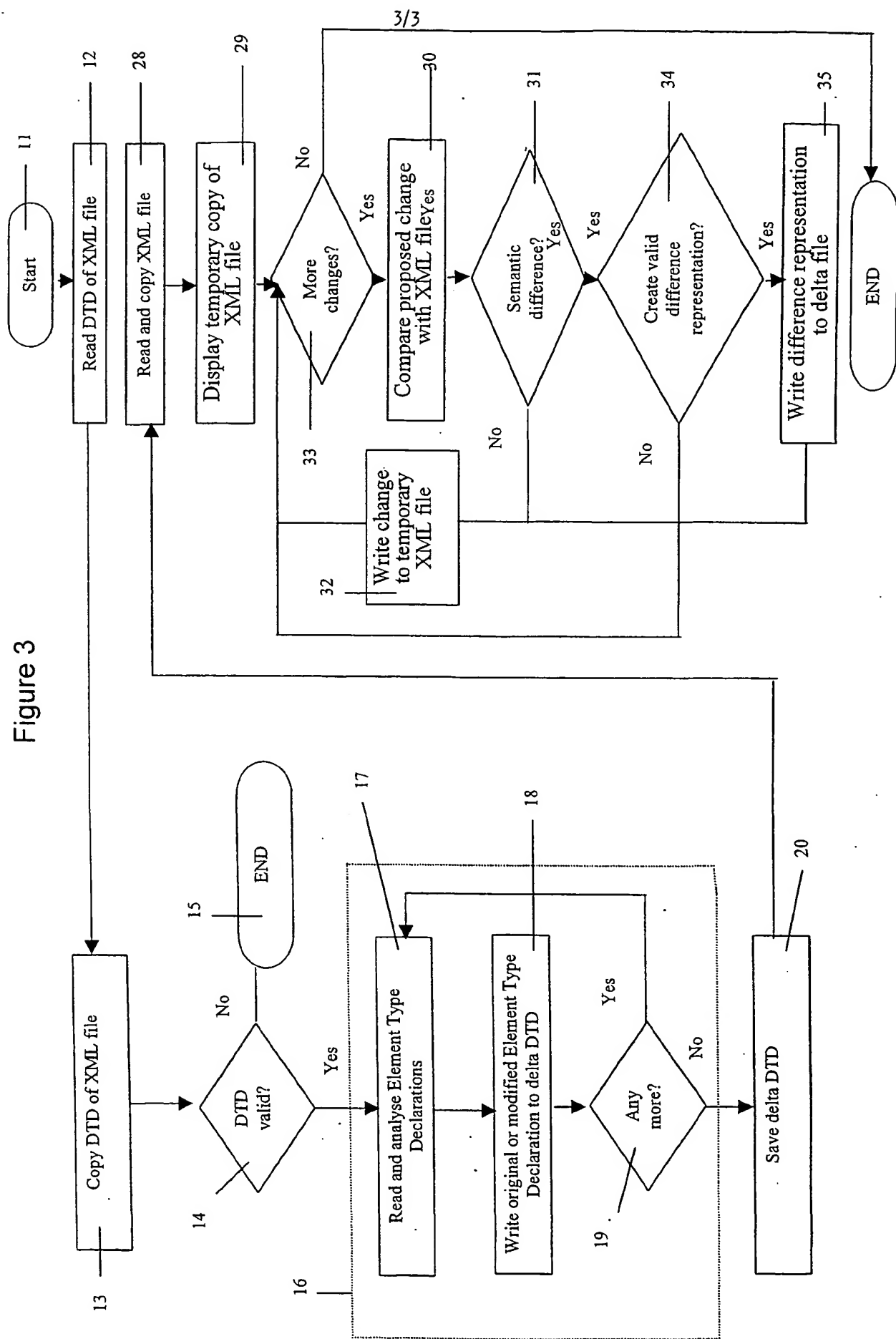




Figure 3





(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 January 2002 (31.01.2002)

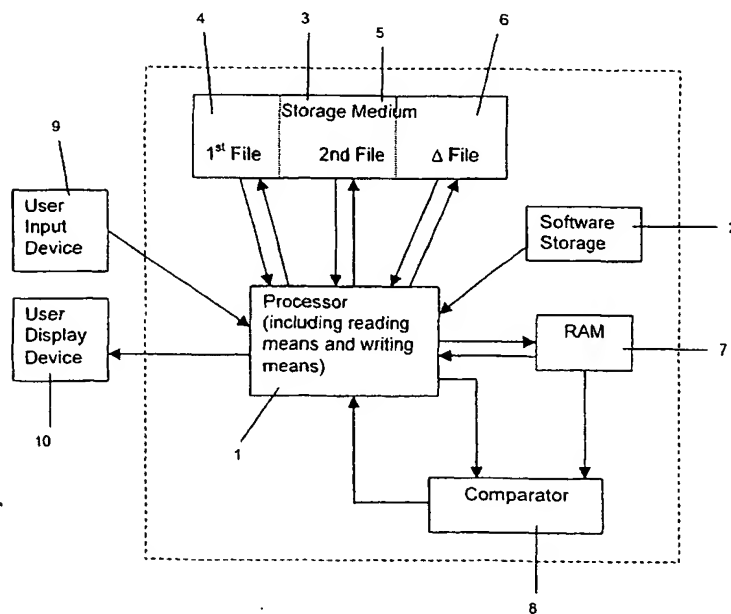
PCT

(10) International Publication Number  
**WO 02/008890 A3**

- (51) International Patent Classification<sup>7</sup>: **G06F 17/30**
- (21) International Application Number: **PCT/GB01/03282**
- (22) International Filing Date: **20 July 2001 (20.07.2001)**
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (30) Priority Data:  
0018042.2                      21 July 2000 (21.07.2000)    **GB**
- (71) Applicant (for all designated States except US): **MONSELL EDM LTD.** [GB/GB]; Monsell House, Monsell Lane, Upton-on-Severn, Worcestershire WR8 0QN (GB).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **LAFONTAINE, Semer, Geoffrey, Thomas** [GB/GB]; Monsell House, Monsell Lane, Upton-on-Severn, Worcestershire WR8 0QN (GB).
- (74) Agents: **PERKINS, Sarah** et al.; Stevens Hewlett & Perkins, Halton House, 20/23 Holborn, London, Greater London EC1N 2JD (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— with international search report

[Continued on next page]

(54) Title: METHOD OF AND SOFTWARE FOR RECORDAL AND VALIDATION OF CHANGES TO MARKUP LANGUAGE FILES



(57) Abstract: A method of recording changes to a markup language file which employs application-defined tags. The changes are recorded in a delta file which is also a markup language file providing validation of the recorded changes against substantially the same markup language structure as that of the markup language file being changed. Where the original markup language file is an XML file with a DTD, a DTD can be created for the delta file which substantially follows the DTD of the original markup language file. Strict compliance of the data recorded in the delta file with the delta DTD provides validation of the changes with respect to the original XML file.

WO 02/008890 A3



(88) Date of publication of the international search report:  
17 April 2003

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, COMPENDEX, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 016 987 A (SUN MICROSYSTEMS INC) 5 July 2000 (2000-07-05) paragraphs '0040!', '0046!'; figures 3-5 abstract	1,2,4-8, 10-16
X	BANGA G ET AL.: "OPTIMISTIC DELTAS FOR WWW LATENCY REDUCTION" PROCEEDINGS OF THE USENIX ANNUAL TECHNICAL CONFERENCE, 1997, pages 289-303, XP000874445 paragraphs '0001!', '03.2!'; figure 6 abstract	1-16
	---	
	--- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

23 October 2002

Date of mailing of the international search report

30/10/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-2016

Authorized officer

Haffner, R

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>CHANG G J S ET AL: "A graphical environment for change detection in structured documents"</p> <p>COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE, 1997. COMPSAC '97. PROCEEDINGS., THE TWENTY-FIRST ANNUAL INTERNATIONAL WASHINGTON, DC, USA 13-15 AUG. 1997, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US,</p> <p>13 August 1997 (1997-08-13), pages 536-541, XP010247357</p> <p>ISBN: 0-8186-8105-5</p> <p>the whole document</p> <p>-----</p>	1-16

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
EP 1016987	A	05-07-2000	US	6377957 B1	23-04-2002
			EP	1016987 A2	05-07-2000
-----					

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**